THE RELATION BETWEEN LANGUAGE LEARNING APTITUDE AND DYNAMIC ASSESSMENT

by

Kristi Ann Moon

A thesis submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Master of Science

in

Speech-Language Pathology

Department of Communication Sciences and Disorders

The University of Utah

December 2014

The University of Utah Graduate School

STATEMENT OF THESIS APPROVAL

The following faculty members served as the supervisory committee chair and members for the thesis of **Kristi Ann Moon**.

Dates at right indicate the members' approval of the thesis.

Sean Redmond, Chair — **8/12/14**
Date Approved

Kathy Chapman, Member — **8/12/14**
Date Approved

Stacy Manwaring, Member — **8/12/14**
Date Approved

The thesis has also been approved by **Michael Blomgren**

Chair of the Department of **Communication Sciences and Disorders**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

The study investigated the relation between dynamic assessment, specifically a morpheme induction task (MIT), and parent ratings of children's language learning aptitude. Dynamic assessment may be an effective tool for testing bilingual children and a positive relationship with parent ratings would help confirm this finding. Secondly, the study aimed to investigate which factors affected performance on the MIT (age, language exposure, vocabulary level, and short-term memory). Twenty-six 5-6-year-old Spanish-English bilingual children and their parents participated. Results found no significant relationship between the MIT and parent ratings of children's language learning aptitude. MIT performance was related to age, language exposure, and short-term memory. Supplemental scoring schemes for the MIT showed different relations among variables. Different error types on the MIT were related to short-term memory, age, and language exposure at various levels. This dynamic assessment was affected by variables besides language abilities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to express my gratitude to all those who made this project possible. Many thanks to Dr. Sean Redmond for his support for my project, his willingness to allow me to research the area I desired, his guidance throughout the project, his funding of the project, for all I learned from our brainstorming sessions, his encouragement during the many speed bumps in my project, and for sharing his abundant knowledge with me. I learned a great deal from him. Thank you to my committee members, Dr. Kathy Chapman and Dr. Stacy Manwaring. I appreciated your input, flexibility, corrections, and caring attitude. Dr. Andrea Ash unofficially helped me throughout my project. I am grateful for her continual support, friendship, and listening ear during the harder days and days of great success. I am indebted for all she's done for me. Also I would like to acknowledge and thank my bilingual assistants Claudia Yazzie, Gisselle Warren, Jessica Carrizo, Ariel Hyams, Joanna Nieves, and Eliza Hintze for sharing their valuable Spanish skills and finding time to call and visit families with me. Additional thanks to Mckenzie Rohde for accompanying me, when working with English speaking families.

I would like to thank the many people involved in the recruitment of participants. Thanks to Dr. Andrea Ash, Jessica Carrizo, and Valerie Green at the University of Utah. Thanks to the principals, secretaries, and others involved at the following schools: Head Start, St. Francis Xavier Catholic School, and Dual Immersion Academy. I appreciate all the families who referred friends and family to my study, as well as all who participated for taking the time participate. I hope they had a good experience and will continue supporting research in the future.

I want to express my gratitude and love to my family. They have supported me through thick and thin even when they thought I was crazy for taking on the challenge of a thesis. Thanks to my mom for making sure I had food to make it through the long days of research activity. I thank my friends for encouraging me to dream big. Finally, I'm grateful for all the prayers answered in relation to my project. It took a lot of people to pull everything together and I'm sure glad they did their part, because this was a great experience for me.

INTRODUCTION

Parental reports are a useful tool for Speech-Language Pathologists (SLP), especially when diagnosing. Parental reports provide important insights into what children do in multiple contexts, across people, and over time. This view is one that SLPs cannot get through mere observation or test administration. In fact, SLPs commonly underestimate children's skills because they often only see children's performance in an unfamiliar environment where language use is artificial (Diamond & Squries, 1993; Hadley & Rice, 1993; Linder, 1993). Parental reports also help parents and SLPs communicate through common terminology (Hadley & Rice, 1993).

Parent Questionnaires (PQs) have been found to be valid for separating cases of typical development (TD) from language impairment (Boudreau, 2005; Dale, 1991; Jackson-Maldonado, et al., 1993; Klee, et al., 1998; Massa, Gomes, Tartter, Wolfson & Halperin, 2008; Rescorla & Alley, 2001; Skarakis-Doyle, Campbell, & Dempsey, 2009). Language impairment (LI) is defined as a "significant language deficit in the face of normal nonverbal intelligence, adequate auditory acuity, and the absence of gross neurological disabilities" (Leonard, Sabbadini, Leonard, & Volterra, 1987, p. 234). Children with LI have semantic and syntactic deficits. Studies have found PQs to be effective when evaluating both vocabulary (Dale, 1991; Rescorla, 1993; Thal, O'Hanlon, Clemmons, & Franlin, 1999) and syntax (Dale, 1991; Thal, O'Hanlon, Clemmons, & Franlin, 1999). Hadley and Rice (1993) used the Speech and Language Assessment Scale

(SLAS), a brief measure of 19 questions, to measure levels of assertiveness, responsiveness, semantic achievement, syntactic proficiency, and articulation accuracy in preschool age children. Results indicated that using the articulation, assertiveness, and semantics composites alone, 85% of children could be correctly classified. Weinberg (1991) analyzed construct validity of the SLAS and found that parent and SLP ratings correlated moderately to highly with standardized tests supporting use of the SLAS as a valid measure. Other PQ measures with high validity include the MacArthur Communicative Development Inventory (Fenson, et al., 1993), classifying 96% of children correctly as TD or LI (Skarakis-Doyle, Campbell, & Dempsey, 2009), the Language Development Survey (Rescorla, 1989, 1993; Rescorla & Alley, 2001) for toddlers, and the Children's Communication Checklist (Bishop, 2003) for school-age children (Norbury, Nash, Baird, & Bishop, 2004). In fact, the Children's Communication Checklist has been shown to separate TD and LI groups better than standardized tests (Bishop, Laws, Adams, & Norbury, 2006; Bishop & McDonald, 2009).

Parent questionnaires have also been used in bilingual populations with promising results. Restrepo (1998) analyzed the effectiveness of an assortment of different tests for school-age bilinguals including language sample analyses, a standardized test, and dynamic measures. Results indicated that combining PQs with a language sample analysis was the most effective way to differentiate cases of typically developing Spanish-English bilinguals from cases of Spanish-English bilinguals with LI. PQ alone achieved specificity of 96% and sensitivity of 74%. Other investigations support the use of PQs as a reliable way of identifying language impairment in bilingual children (Bedore, Pena, Joyner, & Macken, 2011; Gutierrez-Clellen & Kreiter, 2003; Marchman, Martinez-Sussmann, & Dale, 2004; Paradis, Emmerzael, & Duncan, 2010). Spanish PQs include the Spanish Ages and Stages Questionnaire (ASQ) (Squires, Potter, & Bricker, 1999), which was not found to be an effective screener in Spanish (Guiberson, Rodriguez, & Dale, 2011), and Iventarios del Desarrollo de Habilidades Comunicativas (INV) (Jackson-Maldonado, Bates, & Thal, 1992), which is the Spanish Communicative Development Inventory, has been shown to have high concurrent and predictive validity (Jackson-Maldonado, et al., 1993; Thal, Jackson-Maldonado, & Acosta, 2000) for toddlers.

A recent meta-analysis looking at bilingual assessment found the common reference standard when identifying LI to be either parent or teacher report (Dollaghan & Horner, 2011; Gutierrez-Clellen & Pena, 2001; Pena, et al., 2006). Restrepo's (1998) study showed that the PQ measure trumped static and dynamic assessment standardized or otherwise. More support is desired and even required than a valid and reliable PQ in making a diagnosis (Menlove, Gallo, & Arbogast, 2013). Since PQs are currently the most effective measure and other tests are needed, a test that is consistent with PQs should be sought after to assist in diagnosing LI.

Within the United States, the population of bilingual speakers is growing. One out of five children speaks a language other than English (Census 2009). Census records show over 40 languages are spoken by children in U.S. schools with Spanish being the second most spoken language. This means SLPs will be servicing children with increasingly diverse language backgrounds. With this increase of bilinguals comes the growing question of how to diagnose and evaluate which English Language Learners (ELLs) need speech and language services and which simply need more exposure to

English (Bedore & Leonard, 2000; Restrepo, 1998; Saenz & Huer, 2003). Monolingual SLPs will be asked to assess ELL children, and knowledge of how to approach the situation is needed. There has been a movement to create more standardized tests in languages other than English. Spanish tests are being created, but other languages lack resources to assist in diagnosing LI.

With monolingual children, standardized tests along with clinical judgment represent the standard for diagnosing. Standardized tests are static measures that take a snapshot of what children know and compare their performance to same age peers. For monolingual children, it is assumed that they have had sufficient exposure to learn the target (i.e., past tense –ed inflection) but have failed to learn it. Bilingual children that have not had sufficient exposure may not use particular language forms due to lack of experience and not to a deficit in their capacity to learn language. Even when apparently fluent bilinguals are tested they have been found to still test lower on standardized measures than monolingual peers (Cummins, 1984). Paradis, Schneider, and Duncan (2013) looked at vocabulary, narrative, grammatical, and phonologic tests and found that tests standardized on monolinguals ranged in identifying TD bilinguals as LI from 24-78%. They also found the more exposure children had to English the better they performed on static measures.

Another barrier in testing with standardized tests is first language attrition. Language attrition in children prior to 8-10 years of age is considered a shift in dominance to the L2 and incomplete acquisition of L1 (Montrul, 2008). Language attrition is particularly obvious as children are placed in an environment where the L1 is a minority language and exposure becomes limited as children go to school all day in L2

(Schmid, 2011). Anderson (2004) found attrition for Spanish speakers in the U.S. resulted in use of nonspecific vocabulary, loss of gender agreement, and verb morphology issues and perfect tense in Spanish. With those impacts on L1, a typically developing ELL may appear impaired on standardized measures when compared to monolingual peers not experiencing attrition. This compromises the integrity of standardized tests when testing ELL children. Static measures do not appear to be ideal for the ELL student.

Another option when assessing monolingual children is using Dynamic Assessments (DA). DAs are designed to measure children's ability to learn. For children with LI, this may be considered the main deficit as indicated by synonyms for LI such as Language Learning Deficit (Paul & Norbury, 2012). Results would inform a clinician how easy it is to teach a particular child or what kind of scaffolding works best for the child to learn (Vygotsky, 1978). DA is believed to be more independent of language exposure because it evaluates learning aptitude rather than assessing current knowledge. A phonological dynamic assessment measure currently used for screening LI is the Nonword Repetition task (NWR) (Conti-Ramsden, Botting, & Faragher, 2001). For monolingual children, this is thought to be a quick effective method for screening LI (Gray, 2003). Conti-Ramsden and Botting examine NWR and its overall effectiveness in differentiating LI from TD status. Response Operating Characteristics (*r*OC) curves indicated an overall accuracy of 0.90, strongly suggesting that NWR represent an effective method for identifying cases of LI. Campbell, Dollaghan, Needlemen, and Janosky (1997) pointed out that NWR may be preferred over other measures because performance is independent of language knowledge or exposure.

NWR is designed to be culturally unbiased because children are asked to repeat word forms that are not present in their language (Dollaghan & Campbell, 1998). Although NWR may not be as dependent on language exposure as standardized tests, it may still be biased against nonnative speakers due to phonotactics associated with nonword construction. A Welsh-English bilingual study found that inexperienced speakers made more errors on sounds unique to their L2 (Sharp & Gathercole, 2013). In a study by Windsor, Kohnert, Lobutz, and Pham (2010), results of NWR repetition tasks, based on the syllable properties of Spanish and English, with monolingual English speakers and bilingual Spanish-English speakers, found bilinguals scored higher in their dominant language. These findings suggest that NWR may need to be administered in children's dominant language, thus creating the need for language specific NWR tasks. It also indicated that not all NWR tasks are created equal as the two NWR tasks did not have equal sensitivity and specificity. Thordardottir and Brandeker (2013) also found bilinguals to perform below monolinguals on NWR. They found NWR to be effective in differentiating bilingual and monolingual groups with LI and Typical Development (TD), however, results showed bilinguals with TD to score approximate to monolinguals with LI. Separate norms appear to be needed for monolingual and bilingual students for NWR to be an effective way of testing bilinguals.

One reason NWR may be so effective in separating children with LI from TD is that NWR taps into short-term memory (STM). Children with STM deficits will be limited in their ability to store and correctly process phonologic stimuli (Gathercole & Baddeley, 1990). This deficit can explain the difficulties seen in children with SLI such as low vocabulary and poor grammatical marking.

Studies looking at bilingual children have compared NWR performance to vocabulary level (Gathecole, Willis, & Baddeley, 1991; Gorman, 2012). Reports have found a moderate positive association between vocabulary and NWR: $r = 0.4$ to 0.5 (Gathecole, Willis, & Baddeley, 1991; Thordardottir & Brandker, 2013). Gorman (2012) explains this relation may be bidirectional. Children with poor STM may not be able to encode new words as effectively as children with strong phonologic STM, creating a low vocabulary, or children with low vocabularies may not have access to as many phonologic templates to build new words upon and thus their phonologic STM is limited.

While STM and NWR have strong links to language and are effective for monolinguals, they may not be as effective for bilinguals. There seems to be a connection between language specific knowledge (vocabulary, phonotactics) and NWR performance. Thus this may not be the most promising screener for bilingual children (Summers, Bohman, Gillam, Pena, & Bedore, 2010).

Vocabulary learning tasks are another prevalent dynamic assessment type researched. Here the consensus is that the amount of effort needed to teach new vocabulary has been more predictive of LI status than children's pre- and posttest scores on these assessments. Researchers suggest incorporating a "modifiability scale" representing the amount of clinician support children needed to learn new vocabulary as an indicator of language learning impairment (Kapantzoglou, Restrepo, & Thompson, 2012; Patterson, Rodriguez, & Dale, 2013; Pena, Iglesias, & Lidz, 2001; Ukrainetz, Harpell, Walsh, & Coyle, 2000). Modifiability measures have also been found to be reliable when using narratives as the focus of dynamic assessment (Gillam, Pena, & Miller, 1999; Pena, et al., 2006). However, using a modifiability scale is potentially problematic due to its subjective nature. The experience the clinician has had teaching vocabulary will frame their expectations for what was easy and what was hard. Since SLPs work with children who have language impairment, they may see children as typical that are really struggling because of improper expectation. SLPs, especially new SLPs, may be less inclined to use this subjective test and test with inaccurate static measures over switching to a method with which they are not confident using.

Current studies suggest morphosyntax may be a more reliable clinical marker for young children with LI than reduced vocabulary (Bedore & Leronard, 1998; Rice, Wexler, & Hershberger, 1998). If morphosyntax is a more reliable marker for LI, it may also be a more effective venue for DA procedures than vocabulary. While it has been speculated by Leonard (1998) that child speakers with L1 from languages with rich morphological systems would not have as significant of deficits with morphology, some research suggests otherwise. Bedore and Leonard (2000) compared languages with rich and weak morphology systems and found that children from both languages struggled to learn novel words with different inflections. In a different study, children with LI speaking a morphologically rich language were less likely to use and sequence multiple suffixes properly than matched peers (Lukacs, Laurence, & Kas, 2010). Cross-linguistically, morphemes are commonly represented as an area of deficit for children with LI, making morpheme induction tasks potentially useful for diverse sets of bilingual children (Acarlar & Johnston, 2011; Bedore & Leonard, 2005; Gabor & Lukacs, 2012).

Jacobs and Coufal (2001) created a computerized test that screened bilingual children using expressive and receptive vocabulary, expressive morpheme, and verbal sequencing dynamic assessments. They found groups to be separated most efficiently by the expressive morpheme learning task. ELL children with TD had scores that ranged from 94% to 100% and ELL children with LI scores ranged from 6-63% accuracy. They advocated for the procedure because of its group differentiation as well as time effectiveness over other dynamic assessments.

Roseberry and Connell (1991) used an invented derivational morpheme (based on stimuli used in an earlier study by Connell 1987). The invented morpheme indicated “part of”. Schwa was attached as a suffix so “box-ə” meant “part of a box”. Bilingual children were taught the morpheme by contrasting full and partial pictures of common objects along with an image label. Posttesting revealed this task produced sensitivity of 77% and specificity of 92% with 26 participants.

Roseberry and Connell’s nominal morpheme induction task has been extended by other investigators. Boyer and Martin (2012) replicated the study with a modification that compared the task when it was administered in English versus English with a Spanish translation. For the Spanish translation, they used the morpheme /–beI/. They found that language of administration did not affect performance and that the task was independent of participant language knowledge. They did find that the children preferred the morpheme /-beI/ that agreed with Spanish L1 phonotactics and that this probably made the new form more learnable. Children’s age was found to be a factor as 3-4-year-olds were much less likely to learn the morpheme than 5-year-olds. In this case, measures of sensitivity and specificity were not appropriate due to the participants’ young age. In contrast to these results, Kohnert and Danahy (2007) found the task to be greatly influenced by language of administration and that not all children with TD could learn the task in the number of training trials presented (20). They also had a younger age sample but did not find a significant correlation between age and performance. With conflicting data being reported, Kan and Kohnert (2008) criticized Roseberry and Connell for not providing enough demographic information, such as vocabulary level, on the children tested.

Previous investigations suggest that Roseberry and Connell's morpheme induction task holds promise as a language neutral procedure that monolingual SLPs can administer to bilingual children suspected of having LI. However, important gaps in our understanding of the nature of this task limit our enthusiasm for using dynamic assessment. For example, the relation to language exposure, vocabulary level, and age, and the effect of different number of trials is unclear.

Questions

Does the performance of young school-age bilingual children on the morpheme induction task correlate to their language learning aptitude as measured by PQ?

> Null hypothesis: Performance will not be correlated to language learning aptitude.
>
> Alternative hypothesis: Performance will be correlated to language learning aptitude.

How does the performance of young school-age bilingual children on the morpheme induction task correlate to their language exposure, age, vocabulary level, and nonword repetition performance?

> Null hypothesis: Performance will not be correlated to the morpheme induction task.

Alternative hypothesis: Performance will be correlated to the morpheme induction task.

METHODS

Participants

With IRB approval and permission from Dual Immersion Academy, St. Francis Xavier Catholic School, and Head Start, 830 flyers were sent home with preschool, kindergarten, and first-grade students in the Salt Lake Valley. Flyers were also hung on the University of Utah campus. Participating parents were asked if they knew any parents with children who could participate.

Twenty-six Spanish-English bilingual preschool through first-grade students (age range 5;0-6;11) participated in this study. Participants were simultaneous or sequential bilingual learners. All participants passed a hearing screening at 25 dB for 1000, 2000, and 4000 Hz with a calibrated audiometer from the University of Utah Communication Science and Disorders Department. Standard scores for nonverbal cognitive abilities were ≥80 on the Naglieri Nonverbal Ability Test (Naglieri, 2004). Two potential participants were tested but were not included in the study due to scores <80 on the Naglieri Nonverbal Ability Test. Parents reported no learning, hearing, or other disabilities among participants. Three participants were receiving speech services at the time of the study for language delays or speech sounds disorders. All participants had at least 6 months of formal exposure to English. Demographic characteristics associated with the study sample are provided in Tables 1 and 2.

Table 1 Group descriptive statistics.

	M (*SD*) Range
Age (months)	70 (6.84) 60-83
Years in school	1.9 (1.13) .5-5.5
Spanish/English Exposure Ratio	.58 (.10) .31-.80
Spanish/English Proficiency Ratio	.49 (.09) .31-67
Mother's education	2.96 (1.216) 1-5 1 Junior high-5 Master's/ Doctorate degree
Naglieri Nonverbal Ability Test	107.58 (13.21) 86-133
Nonword Repetition-English	75.40 (10.61) 59-94
Nonword Repetition –Spanish	80.27 (12.58) 35-94
Receptive One Word Picture Vocabulary Test: Spanish Bilingual Edition- standard score	103.36 (8.34) 90-123

Table 2 Additional group statistics.

	#		%	
Female	16		61.5	
Native language				
Spanish	22		84.6	
English	2		7.7	
Both	2		7.7	
Proficient language				
Spanish	13		50	
English	8		30.8	
Both	5		19.2	
Family's Country of Origin	Mother	Father	Mother	Father
Mexico	16	17	61.5	65.4
Guatemala	2	2	7.7	7.7
Peru	2	3	7.7	11.5
Colombia	2	1	7.7	3.8
Honduras	1	0	3.8	0
US	3	2	11.5	7.7
# of Siblings				
0	4		15.4	
1-2	12		46.2	
3+	10		38.5	
Birth Order				
1	12		46.2	
2	6		23.1	
3	5		19.2	
4+	3		11.5	
Speech Services	3		11.5	
Grade in School				
Pre-K	11		42.3	
Kindergarten	9		34.6	
1st grade	6		23.1	
Free lunch program	16 (*N* 22)		61.5	
Participant schools				
Dual Immersion Academy	4		15.4	
Head Start	9		34.6	
St. Francis Xavier	2		7.7	
Catholic				
school	4		15.4	
St. John the Baptist school	4		15.4	
Other public schools	3		11.5	
Other private schools				

Procedures

Four different tasks were administered to the participants: the *Receptive One-Word Picture Vocabulary Test, 2012 Edition- Spanish-Bilingual Edition* (*ROWPVT: SBE*) (Brownell, 2012), a standardized vocabulary tests for bilinguals, nonword repetition tasks in Spanish (Ebert, Kalanek, Cordero, & Kohnert, 2008) and English (Dollaghan & Campbell, 1998), and the derivational morpheme induction task (Roseberry & Connell, 1991). The parent questionnaires were completed by participant parents or guardians. Parent questionnaires included demographic information, the *Speech and Language Assessment Scale* (Hadley & Rice, 1993), and the *Ability and Use of Language in the Home* (Gutierrez-Clellen, Kreiter, & Zagursky, 2007). The protocol order varied across participants to control for order effects.

Before beginning the test procedure or collecting information from parents, individuals were informed of their right to not participate. It was explained that all information would be kept confidential and locked at the University of Utah language acquisition lab. To protect the identity of all participants, an alphanumeric code was used on all score sheets. All parents were asked whether they would like a Spanish interpreter on site to ensure participants could express needs, wants, and allow them to excuse themselves. Interpreters were present unless the families declined these services.

Parent questionnaire

Three parent questionnaires (PQs) were used to gather information from parents about each participant’s background, language exposure, and language learning aptitude. PQs were explained by a trained researcher to the parent or grandparent in the language in which they expressed feeling most comfortable. Language learning aptitude was measured by a translated copy of the *Speech and Language Assessment Scale* (SLAS) developed by Hadley and Rice (1993). Parents or grandparents answered questions relating to both participants' Spanish and English language abilities compared to other children the same age for five domains of language (Semantics, Assertiveness, Responsiveness, Syntax, and Articulation). Comparing their child to other children, the same age helped add an element of standardization. Hadley and Rice found parents and SLPs to be equally reliable rating the language abilities of children using the SLAS. Four individuals, including parents, teachers, and SLPS rated each child. Only 13 of the 19 questions were reliably answered by all four raters. These 13 questions were compared to standard measures such as mean length of length, Peabody Picture Vocabulary Test (Dunn & Dunn, 1981), and the Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 1986) and was found to have moderate-moderately high construct validity (Weinberg, 1991). The 13 reliable questions were asked using a 7-point Leikart scale. Information was gathered over the phone or in person by a trained researcher. Language exposure was measured using the *Ability and Use of Language in the Home* questionnaire (Gutierrez-Clellen, Kreiter, & Zagursky, 2007). The language exposure questionnaire asked parents to rate how much Spanish and English participants hear from various family members at home on a 4-point scale.

<u>Receptive vocabulary task</u>

The *Receptive One-Word Picture Vocabulary Test, 2012 Edition- Spanish-Bilingual Edition* (*ROWPVT: SBE*) (Brownell, 2012) tests an individual's receptive vocabulary through the use of picture identification. The test gives prompts in both Spanish and English and measures the individual's combined vocabulary knowledge.

Testing norms were established by testing bilinguals of various proficiencies in both languages. To be consistent with other measures given, the test was computerized and had prerecorded audio for instructions and prompts in both Spanish and English. Basal and ceiling rules were followed. To keep administration standardized across participants, the prompt was always given in both languages.

Nonword repetition tasks

A Spanish and English NWR task were administered. Both tests were designed to follow each language's phonotactics but not to mimic specific words within the language. The Spanish task was developed by Ebert, Kalanek, Cordero, and Kohnert (2008). It contained 120 phonemes in 20 words ranging from one to five syllables. The English task was developed by Dollaghan and Campbell (1998). It contained 96 phonemes with word ranging from one to four syllables. The inclusion of five syllable words for Spanish was found to be more appropriate due to the frequency of five syllable words in Spanish as compared to English (Ebert, Kalanek, Cordero, & Kohnert, 2008). Students listened to the words via headphones and a MP3 player. Words were recorded by a native speaker. Students were instructed to repeat each nonsense word.

Morpheme induction task

Procedures for the morpheme induction task (MIT) were modifications of those used by Roseberry and Connell (1991). The Spanish version (Boyer & Martin, 2012; Kohnert & Danahy, 2007) of the morpheme (/-beI/) used by Connell (1987) was taught during one 10-15-minute session. Stimuli were presented to each participant using PowerPoint on a laptop display. Directions to the task were given in Spanish and English, the training items were only presented in English. A vocabulary set of 15 pictures was presented first and participants were directed to label the pictures. If the participant did not know the word, it was provided for them. This was a pretest of words used in the generalization probes following the task. The induction task was then introduced by telling participants that they were going to hear a funny lady talk. They were then told to listen and not to speak. Participants were directed to look at a computer display showing a single picture. Recorded audio either said “This is a NOUN, a NOUN” while showing the whole image, or a partial image was displayed and the recording said “This is a NOUN-/beI/, a NOUN-/beI/”. Picture sets were kept together but order varied whether the complete or partial image was shown first. After all 20 picture pairs were presented, a generalization probe was administered. Ten images from the vocabulary portion were presented and their contrast; sets were kept together but order varied. Participants were instructed to name the pictures speaking like the funny lady. The next picture was presented after a response was given. The responses were recorded for later scoring as well as to indicate whether phase II of teaching was required. If participants did not reach 85% accuracy on the generalization probe, then phase II was administered. Phase II consisted of 10 more teaching sets followed by a second generalization probe of five items in the same manner as above.

Scoring

Parent questionnaires

Three parent questionnaires were used in this study. First, demographic information was entered with numeric codes into SPSS for analysis. Second, the SLAS ratings between 1 and 7 were entered into SPSS for each language. Hadley and Rice

(1993) combined the 13 items into five composite areas of language. Scores within each area were averaged to create assertiveness, responsiveness, syntax, articulation, and semantics composite scores for Spanish and for English. Third, language exposure and language proficiency ratios were calculated. A study by Bedore, et al. (2012) suggested the most effective measurement of language exposure is the child's current language input. Parents rated on a scale from 1 (never) – 4 (always) how much Spanish and English they heard from the mother, father, and others in the home. These three scores were averaged for both languages. The Spanish score was then divided by the combined Spanish and English score to create a Spanish to English exposure ratio. Higher values meant more Spanish was spoken at home than English. For the proficiency score, parents rated the participant's speaking and understanding of Spanish and English on a scale of 1 (doesn't speak/no understanding) – 5 (always speaks/understands everything). Understanding and speaking were averaged for both languages. The proficiency ratio was calculated the same way as the exposure ratio with Spanish being divided by Spanish and English. Higher scores meant they were more proficient in Spanish than in English.

Nonword repetition tasks

This task was recorded and relistened to before scoring nonwords phoneme by phoneme. Each task was listened to by a native speaker of the language. Percentage of correct phonemes was used in data comparisons. A score was calculated for English and Spanish. To account for different language dominances among participants but still have a collective way to describe their STM skills, their higher score was selected as their "best" score. All three scores (Spanish, English and Best) are used in analyses.

Vocabulary

Standard scoring procedures were followed for the Receptive One-Word Picture Vocabulary Test: Spanish-Bilingual Edition (Brownell, 2012). Traditional administration of the test would be administered in the child's dominant language with prompts in their less proficient language when they hesitated. For this study, words were always presented in both languages. Norms were used despite the change in administration. With a 2-year age range between participants, mean scores varied greatly and raw scores were not appropriate for comparisons.

Morpheme induction task

After training items were presented to teach the new morpheme, 10 picture sets (a complete picture and partial picture) were administered to each participant allowing for 20 possible responses. Words missed during the vocabulary pretest were not included for scoring unless the participant demonstrated knowledge of the word spontaneously during the generalization probe. Responses such as "I don't know" or "I don't want to answer" were also not scored. Because of this variability, percentages were used to compare participants' scores. Words intended to be unmarked were scored as correct, or incorrect (if marked, i.e., ball-beI). Words intended to be marked were scored as correct or incorrect (for morpheme omission, altered morphemes, or phrasal modifications). A combined percentage correct for both marked and unmarked targets was used in the total percent correct (%correct). For participants who were administered the second set of training item, a separate %correct score was calculated for the five sets of generalization items.

RESULTS

Complete data were available for 23 of the 26 participants. Missing data were due to recording equipment failure (1 participant) and to instances of participant non-responses to some items on the vocabulary measure or nonword repetition tasks (2 participants).

Question 1

Does bilingual performance on the morpheme induction task correlate to level of language learning aptitude as measured by PQ?

Ten composites, five for each language, were created from the parent questionnaire in regards to language learning aptitude. To answer the first research question related to the association between morpheme induction task (MIT) performance and parent ratings, as measured by the Speech and Language Assessment Scale (SLAS), the feasibility of data reduction was considered first. Pearson product correlations were run to identify the degree of separation among these composites for each language. Associations among Spanish and English SLAS composites were also examined. The findings in Table 3 indicated that within each language, the five composites were highly correlated with one another. All five Spanish domains correlated with one another with $r>.840$, $p<.001$ and all English domains correlated with $r>.868$, $p<.001$. This indicated that within languages, parents rated participants similarly across the different language areas rather than rating one domain higher or lower than another (i.e., all 5s and 6s, or all

Table 3. Zero order Pearson product correlations among Speech and Language Assessment Scale (SLAS) Composites: English and Spanish.

		Spanish Assert	Spanish Semantics	Spanish Artic	Spanish Response	Spanish Syntax	English Assert	English Semantics	English Artic	English Response	English Syntax
Spanish Assert	Corr.	-									
	Sig.										
Spanish Semantics	Corr.	.930**	-								
	Sig.	.000									
Spanish Artic	Corr.	.915**	.890**	-							
	Sig.	.000	.000								
Spanish Response	Corr.	.919**	.843**	.880**	-						
	Sig.	.000	.000	.000							
Spanish Syntax	Corr.	.922**	.907**	.907**	.859**	-					
	Sig.	.000	.000	.000	.000						
English Assert	Corr.	.123	.101	.126	.193	.125	-				
	Sig.	.550	.623	.541	.344	.542					
English Semantics	Corr.	.356	.384	.380	.409*	.375	.895**	-			
	Sig.	.074	.053	.056	.038	.059	.000				
English Artic	Corr.	.390*	.391*	.470*	.413*	.415*	.879**	.903**	-		
	Sig.	.049	.048	.016	.036	.035	.000	.000			
English Response	Corr.	.175	.135	.196	.256	.169	.952**	.889**	.868**	-	
	Sig.	.392	.511	.337	.206	.408	.000	.000	.000		
English Syntax	Corr.	.319	.335	.404*	.359	.372	.872**	.904**	.950**	.881**	-
	Sig.	.112	.094	.041	.072	.061	.000	.000	.000	.000	

* $p < .05$, ** $p < .01$
2 tailed
N 26
Assert=Assertiveness; Artic= Articulation; Response=Responsiveness

4s). Levels of association between Spanish and English composites ranged from small to moderate (*r* range: .101 to .470), suggesting that parents in some areas were providing differentiation in their estimates between languages (i.e., 5s for Spanish but 4s for English). To the extent that composites display high levels of intercorrelation within a language, it would not be necessary to include all of the composites in subsequent stages of the analysis. Outcomes supported the use of a single SLAS composite score for each language. Given that some children may have mixed dominance between Spanish and English, each item was reviewed and the higher score for each item was used to calculate an average rating. This third average is referred to as the "Best Composite" score in future analyses.

After data reduction, Pearson product correlations were run to look for associations among the MIT %correct, the Spanish SLAS Composite, English SLAS Composite, and Best SLAS Composite scores. Results are displayed in Table 4 and indicate that parent ratings of language learning ability were not significantly correlated with MIT performance for Spanish (r= -.100, p=.627), English (r= .273, p=.177), or Best (r= .060, p=.773) SLAS composites.

Results for question 1 indicated that parent ratings of language learning ability did not correlate with children's performance on the morpheme induction task and therefore, the null hypothesis that there are no associations among these variables could not be rejected.

Table 4. Zero order Pearson product correlation among Speech and Language Assessment Scale (SLAS) composites, and morpheme induction task %correct

	SLAS English composite	SLAS Spanish composite	SLAS Best composite	MIT %correct
SLAS English composite Corr. Sig.	-			
SLAS Spanish composite Corr. Sig.	.323 .108	-		
SLAS Best composite Corr. Sig.	.670** .000	.744** .000	-	
MIT %correct Corr. Sig.	.273 .177	-.100 .627	.060 .773	-
**p*<.05, ***p*<.01 2 tailed *N* 26				

Question 2

How does bilingual performance on the morpheme induction task correlate to language exposure, age, vocabulary level, and nonword repetition performance?

To address Question 2, which examined if children's MIT performance was associated with their vocabulary level, language exposure, age, or STM rather than parental ratings, Pearson product correlations were run among MIT %correct, vocabulary standard scores, a Spanish/English exposure ratio, age, and NWR measures. See Table 5 for results. Results indicated that participants' receptive Spanish-English vocabulary was not significantly associated with their MIT performance (r= .192, p=.357, small effect size). Language exposure, as measured by the Spanish/English Exposure Ratio, was calculated by dividing reported levels of input in Spanish from mother, father, and other family members by levels of the Spanish input and English input. Higher values would indicate a ratio favoring Spanish input over English input. This ratio of language

Table 5. Zero order Pearson product correlation among NWR tasks, ROWPVT:SBE standard scores, language exposure, age, and morpheme induction task %correct.

		NWR English	NWR Spanish	NWR Best	ROWPVT Standard score	Language Exposure Ratio	Age	MIT %correct
NWR English	Corr. Sig. *N*	-						
NWR Spanish	Corr. Sig. *N*	.685** .000 25	-					
NWR Best	Corr. Sig. *N*	.759** .000 25	.981** .000 25	-				
ROWPVT Standard score	Corr. Sig. *N*	.384 .064 24	-.055 .799 24	.0002 .992 24	-			
Language Exposure Ratio	Corr. Sig. *N*	-.295 .152 25	.253 .223 25	.195 .349 25	-.413* .040 25	-		
Age	Corr. Sig. *N*	.449* .024 25	.138 .510 25	.234 .259 25	.095 .651 25	-.240 .237 26	-	
MIT %correct	Corr. Sig. *N*	.584** .002 25	.350 .087 25	.357 .080 25	.192 .357 25	-.444* .023 26	.436* .026 26	-

*p<.05, **p<.01
2 tailed
NWR=Nonword repetition; ROWPVT= Receptive One-word Picture Vocabulary Test: Spanish-Bilingual Edition, 2012; MIT= morpheme induction task

exposure was significantly negatively correlated with participants' MIT performance (r= -.444, p=.023) with a moderate effect size, meaning MIT scores increased as English input increased. Chronological age was also significantly correlated with MIT performance (r= .436, p=.026) with a moderate effect size. Short-term memory (STM) was measured by NWR performance. Associations between participant's performance on the English NWR, Spanish NWR, and Best score between English and Spanish show that English NWR was significantly correlated with MIT performance (r= .584, p=.002) with a large effect size but Spanish NWR (*r*= .350, *p*=.087) and Best NWR (*r*=.357, *p*=.080) both had moderate effects sizes but were not correlated with MIT performance. Most participants (20/25) scored higher in Spanish and therefore, Spanish and Best NWR performance were quite similar.

To understand the relative impact each of these variables had on participant's MIT performance, a standard multiple regression analysis was used to determine the size of the overall relationship between the dependent MIT performance (predicted) and the independent (predictor) variables of STM (English NWR), language exposure, and age. The model (Table 6) was significant (F (3,21) =5.011, p=.009) with English NWR (p=.028) significantly predicting MIT performance. The regression formula was %correct= 14.892 + -.257(exposure) + .122(age) + .453(NWR-E). These variables explained 41.7% of the variability in participants' MIT %correct scores.

Results associated with question 2 indicated that language exposure, age, and STM were associated with participants' MIT performance, with English NWR alone accounting for a moderate portion of the variation among scores. These findings provide partial support to the hypothesis that knowledge-independent language processing variables, such as NWR, would be associated with MIT performance.

Table 6. Standard multiple regression for MIT %correct, language exposure, age, and English NWR

	Beta weight	Beta	*p*-value	R^2	F	*p*-value
Intercept	14.892		.767	.417	5.011	.009
Language exposure ratio	-53.026	-.257	.156			
Age (mos)	.391	.122	.521			
NWR-English	.910	.453	.028*			
MIT= Morpheme induction task; NWR= Nonword repetition						

Supplemental Analyses

Associations among MIT measure and other demographic/developmental variables

Dynamic assessment is theoretically designed to circumscribe factors such as language exposure and create an equal learning environment for language learners. To examine these claims more closely, additional factors were analyzed to see if they also had an impact on MIT performance. Pearson production correlations were run to look for associations among MIT %correct, nonverbal abilities, years in school, language proficiency, mother's education level, birth order, and number of siblings. Results of these correlations are displayed in Table 7. Nonverbal abilities (r= -.075, p=.717), years in school (r= .179, p=.382, small effect size), mother's education level (r= -.006, p=.978), birth order (r= -.042, p=.838), and number of siblings (r= -.168, p=.412, small effect size) were not significantly correlated. Participants' understanding and speaking skills in Spanish and English were rated by parents to create the proficiency ratio. The Spanish/English proficiency ratio was calculated by adding speaking and understanding skills in Spanish, then that was divided by the total for understanding and speaking of both Spanish and English. The higher the number, the more proficient the participant was in Spanish compared to English. The correlation between proficiency and MIT performance was not significant (r= -.336, p=.093) but had a medium effect size. None of these additional factors contributed to children's MIT performance.

Typology of Errors Produced

Errors represented 19.6% (95/484) of the responses provided by the participants on the morpheme induction task. For this reason, the influence of error types was

Table 7. Zero order Pearson product correlations among MIT %correct and other variables.

		Nonverbal abilities	Years in school	Spanish: English Proficiency ratio	Mother's education	Birth order	# of siblings	MIT %Correct
Nonverbal abilities	Corr.	-						
	Sig.							
Years in school	Corr.	.336	-					
	Sig.	.093						
Spanish:English Proficiency ratio	Corr.	-.220	-.319	-				
	Sig.	.280	.112					
Mother's education	Corr.	.136	.230	.169	-			
	Sig.	.508	.258	.411				
Birth order	Corr.	.078	-.136	-.039	-.197	-		
	Sig.	.705	.507	.849	.334			
# of siblings	Corr.	-.147	-.233	.519**	-.056	.641**	-	
	Sig.	.473	.253	.007	.788	.000		
MIT %Correct	Corr.	-.075	.179	-.336	-.006	-.042	-.168	-
	Sig.	.717	.382	.093	.978	.838	.412	

**p*<.05, ***p*<.01
2 tailed
N 26

considered. Errors fit into four general categories: false alarms, omissions, commission errors, and phrasal modifications. Examples of each error type are given in Table 8. False alarms represented productions of the new morpheme in untrained contexts. For example, labeling the picture depicting the entire image of a ball as "*ball-beI"* was a false alarm. Commission errors consisted of phonetically altering the morpheme in the trained context or changing the position of the morpheme (using it as a prefix rather than suffix). Figure 1 and Table 9 display the frequency of these different errors. Commission errors were of main interest given that the production of these types of errors indicated participants understood some aspects of the task but could not execute it accurately. Additional metrics were designed to capture different degrees of "partially correct" responses. Statistical tests were rerun to see if the new metrics would change the outcome to the research questions.

Question 1 asked if MIT performance was correlated with parent ratings. Pearson product correlations were run with Spanish, English, and Best SLAS composites and two additional outcome measures for the morpheme induction task: %suffixed and %morphemed. These new metrics counted specific commission errors as correct.

Table 8. Description of error types.

Error Description	**Target**	**Example(s)**
False alarm	Ball	*"Ball-beI"*
Omission	Ball-beI	*"Ball"*
Comission- Incorrect suffix	Ball-beI	*Ball-vee*
Comission- Prefix error	Ball-beI	*"beI-ball", "mo-ball"*
Phrasal modification	Ball-beI	*"Half of a ball", "Little ball"*

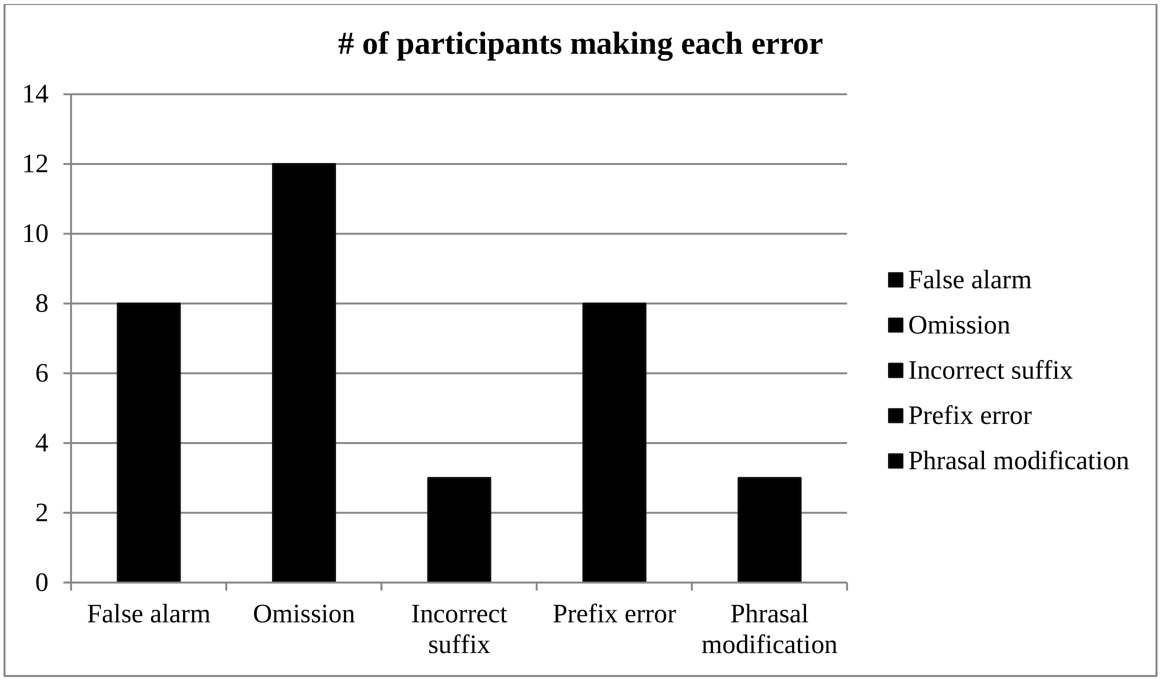

Figure 1. Frequency of participants making each error type.

Table 9. Frequency of error types by number of responses.

	Target- marked		Target-unmarked
Response- marked # of responses % of responses	Hit 160 33.1%		False Alarm 15 3.1%
Response- unmarked # of responses % of responses	Miss 80 16.5%	31 Omission 30 Commission 19 Phrasal __ Modification 80	Correct Non-use 229 47.3%
Hit+ False Alarm+ Miss+ Correct Nonuse=100% Ideal outcome is 50% Hits and 50% Correct Nonuses 36 responses with root errors were not scored			

Responses included as correct for %suffixed were target responses and those responses that contained a phonetically altered suffix (i.e., *"ball-bo*"). Responses included as correct for %morphemed were target responses and any commission error (i.e., *ball-bo, beI-ball, bo-ball*). %Morphemed is similar to %suffixed but also credited off-target prefix errors. Results are displayed in Table 10 and show that %suffixed was not significantly correlated with parent ratings of language learning ability for the Spanish (*r*= -.037, *p*=.858), English (*r*= .247, *p*=.224), or Best (*r*= .097, *p*=.639) SLAS composites. Additionally, %morphemed, shown in Table 10, was not significantly correlated with parent ratings for Spanish (*r*= -.026, *p*=.898), English (*r*= .260, *p*=.200), or Best (*r*= .091, *p*=.658) SLAS composites. Accounting for these different error types did not impact general conclusions for Question 1. MIT performance for any MIT outcome measure was not correlated with parent ratings.

Question 2 considered if MIT performance was impacted by vocabulary, language exposure, age, and STM. Pearson product correlations were run with the MIT %suffixed, MIT %morphemed, vocabulary level, language exposure ratio, age, and NWR performances. Table 11 displays the results, which indicated that %suffixed was not significantly correlated with Spanish-English vocabulary standard scores (*r*= .081, *p*=.700) or language exposure (*r*= -.289, *p*=.153, small effect size) but it was significantly associated with age (*r*= .560, *p*=.003) with a large effect size and STM for all three NWR scores: English (*r*= .554, *p*=.004, large effect size), Spanish (*r*= .465, *p*=.019, moderate effect size), and Best (*r*= .464, *p*=.020, moderate effect size).

Table 10. Zero order Pearson product correlation among Speech and Language Assessment Scale (SLAS) composites and morpheme induction task (MIT) for %suffixed and %morphemed.

		SLAS English composite	SLAS Spanish composite	SLAS Best composite	MIT %suffixed	MIT %morphemed
SLAS English composite	Corr. Sig.	-				
SLAS Spanish composite	Corr. Sig.	.323 .108	-			
SLAS Best composite	Corr. Sig.	.670** .000	.744** .000	-		
MIT %suffixed	Corr. Sig.	.247 .224	-.037 .858	.097 .639	-	
MIT %morphemed	Corr. Sig.	.260 .200	-.026 .898	.091 .658	.955** .000	-
*p<.05, **p<.01 2 tailed *N= 26*						

Table 11. Zero order Pearson product correlation among NWR tasks, ROWPVT:SBE standard scores, language exposure, age, and morpheme induction task %suffixed.

		NWR English	NWR Spanish	NWR Best	ROWPVT Standard scores	Exposure Ratio	Language Age	MIT %suffixed
NWR English	Corr. Sig. *N*	-						
NWR Spanish	Corr. Sig. *N*	.685** .000 25	-					
NWR Best	Corr. Sig. *N*	.759** .000 25	.981** .000 25	-				
ROWPVT Standard scores	Corr. Sig. *N*	.384 .064 24	-.055 .799 24	.002 .992 24	-			
Language Exposure Ratio	Corr. Sig. *N*	-.295 .152 25	.253 .223 25	.195 .349 25	-.413* .040 25	-		
Age	Corr. Sig. *N*	.449* .024 25	.138 .510 25	.234 .259 25	.095 .651 25	-.240 .237 26	-	
MIT %suffixed	Corr. Sig. *N*	.554** .004 25	.465* .019 25	.464* .020 25	.081 .700 25	-.289 .153 26	.560** .003 26	-

*$p<.05$, **$p<.01$
2 tailed
NWR=Nonword repetition; ROWPVT= Receptive One-word Picture Vocabulary Test: Spanish-Bilingual Edition, 2012; MIT= morpheme induction task

The findings for %morphemed, displayed in Table 12, were similar. %Morphemed was not significantly correlated with vocabulary level (*r*= -.029, *p*=.889) or language exposure (*r*= -.168, *p*=.411, small effect size) but was significantly correlated with age (*r*= .526, *p*=.006) with a large effect size, and STM with a moderate effect size for the three NWR scores: English (*r*= .467, *p*=.019), Spanish (*r*= .469, *p*=.018), and Best (*r*= .452, *p*=.023).

Table 12. Zero order Pearson product correlation among NWR tasks, ROWPVT:SBE standard scores, language exposure, age, and morpheme induction task %morphemed.

		NWR English	NWR Spanish	NWR Best	ROWPVT Standard score	Language Exposure Ratio	Age	MIT %morph.
NWR English	Corr. Sig. *N*	-						
NWR Spanish	Corr. Sig. *N*	.685** .000 25	-					
NWR Best	Corr. Sig. *N*	.759** .000 25	.981** .000 25	-				
ROWPVT Standard score	Corr. Sig. *N*	.384 .064 24	-.055 .799 24	.002 .992 24	-			
Language Exposure Ratio	Corr. Sig. *N*	-.295 .152 25	.253 .223 25	.195 .349 25	-.413* .040 25	-		
Age	Corr. Sig. *N*	.449* .024 25	.138 .510 25	.234 .259 25	.095 .651 25	-.240 .237 26	-	
MIT %morph.	Corr. Sig. *N*	.467* .019 25	.469* .018 25	.452* .023 25	-.029 .889 25	-.168 .411 26	.526** .006 26	-

**p*<.05, ***p*<.01
2 tailed
NWR=Nonword repetition; ROWPVT= Receptive One-word Picture Vocabulary Test: Spanish-Bilingual Edition, 2012; MIT %morph.= morpheme induction task %morphemed

It appears that differentiating errors by type changed some of the associations among MIT and the other variables. When phonetically altered suffix and prefix errors were considered as correct, language exposure was no longer associated with MIT performance and all STM tasks became associated with MIT performance. Vocabulary and age were not impacted by errors types. Vocabulary level remained unassociated with MIT performance. Age was consistently significantly correlated with MIT performance for all scoring schemes.

Because of these changes in significance, new regressions were analyzed based on the new association levels between %suffixed and %morphemed and their significantly correlated variables. A standard multiple regression analysis was used to determine the size of the overall relationship between the dependent MIT %suffixed (predicted) measure and the independent (predictor) variables of Best NWR, English NWR, Spanish NWR, and age. The results (Table 13) yielded different outcomes than when %correct was considered. Findings for %suffixed were significant (F (4,20) = 5.727, p=.003) with age (p=.013), Spanish NWR (p=.030), and Best NWR (p=.048) all significantly contributing to the model. The formula was %suffixed= -64.322 + .506(age) + .467(NWR-E) + 2.249(NWR-S) + -2.216(NWR-B). This model accounts for 53.4% of the variance in %suffixed scores. This is an increase over 41.7% of variance accounted for in the %correct model.

A standard multiple regression analysis was used to determine the size of the overall relationship between the dependent MIT %morphemed (predicted) measure and the independent (predictor) variables of Best NWR, English NWR, Spanish NWR, and

Table 13. Standard multiple regression for MIT %suffixed, age, and NWR.

	Beta weight	Beta	*p*-value	R^2	*F*	*p*-value
Intercept	-64.322		.145	.534	5.727	.003
Age (mos.)	1.498	.506	.013			
NWR English	.868	.467	.105			
NWR-Spanish	5.082	2.249	.030			
NWR-Best	-5.316	-2.216	.048			
MIT= Morpheme induction task; NWR= Nonword repetition						

age. The regression analysis for %morphemed yielded yet different results. Results (Table 14) were significant (F (4,20) = 2.887, p=.049) with age (p=.039) being the only significant predictive variable in the model, %morphemed= -164.361 + .487(age) + .224(NWR-E) + 1.551(NWR-S) + -1.406(NWR-B). This model accounted for 36.6% of variance in the%morphemed score, which was less than the percent variance accounted for in the %correct and %suffixed regression models.

Regression models varied between different MIT outcome measures. Initially, only English NWR predicted the %correct; however, English NWR was no longer a contributor when %suffixed was analyzed. Instead STM scores and age became factors in predicting the variance. However, STM was no longer a contributor for %morphemed, leaving %morphemed to be only significantly predicted by age. MIT performance can be more fully understood when the different error types are analyzed with STM and age as significant variables.

The supplemental analysis that investigated contributions from additional demographic and developmental variables was again examined with the new MIT scoring schemes. Pearson product correlations were run among nonverbal abilities, years in school, Spanish/English proficiency ratio, mother's education, birth order, number of

Table 14. Standard multiple regression for MIT %morphemed, age, and NWR.

	Beta weight	Beta	*p*-value	R^2	*F*	*p*-value
Intercept	-164.361		.072	.366	2.887	.049
Age (mos.)	2.482	.478	.039			
NWR-English	.729	.224	.493			
NWR-Spanish	6.143	1.551	.181			
NWR-Best	-5.914	-1.406	.266			
MIT= Morpheme induction task; NWR= Nonword repetition						

siblings, MIT %suffixed, and MIT %morphemed. None of the variables were significantly associated with MIT performance. Results (Table 15) did not change.

<u>Consequences of Increasing Number of Training Items</u>

Previous investigations have varied with regard to the number of training items used. Results across studies may have varied for this reason. The potential benefit of additional training items was considered. In this study, if an 85% proficiency level was not achieved after the first 20 training items, then 10 additional training items were administered. Eleven of the 26 (42.3%) participants did not meet proficiency and had additional training items administered. A dependent *t*-test analysis compared MIT %correct scores after training 1 and training 2 for the 11 participants who received additional training items. Results were not significant between scores for training 1 and training 2 (M_1= 55.64, M_2= 58.45, t= -.409, p=.691). Because this *t*-test may have been underpowered, MIT %correct scores for all 26 participants were compared. If participants were not given additional training items, they received the same score for training 1 and training 2. This dependent *t*-test compared MIT %correct scores for training 1 and training 2 for all 26 participants. Results (M_1= 79.12, M_2= 80.31, t= -.418, p=.679) were not significant even though additional subjects were included in the analysis.

Table 15. Zero order Pearson product correlations among morpheme induction task (MIT) %suffixed, %morphemed, and other variables.

		Nonverbal abilities	Years in school	Spanish: English Proficiency ratio	Mother's education	Birth order	# of siblings	MIT %Suffixed	MIT %Morphemed
Nonverbal abilities	Corr.	-							
	Sig.								
Years in school	Corr.	.336	-						
	Sig.	.093							
Spanish:English Proficiency ratio	Corr.	-.220	-.319	-					
	Sig.	.280	.112						
Mother's education	Corr.	.136	.230	.169	-				
	Sig.	.508	.258	.411					
Birth order	Corr.	.078	-.136	-.039	-.197	-			
	Sig.	.705	.507	.849	.334				
# of siblings	Corr.	-.147	-.233	.519**	-.056	.641**	-		
	Sig.	.473	.253	.007	.788	.000			
MIT %Suffixed	Corr.	-.041	.240	-.264	-.046	-.071	-.046	-	
	Sig.	.843	.238	.193	.852	.730	.822		
MIT %Morphemed	Corr.	-.095	.253	-.213	-.103	.003	.079	.955*	-
	Sig.	.664	.212	.295	.616	.987	.702	.000	

*p<.05, **p<.01
2 tailed
N= 26

Thus, at the group level, additional practice items appeared to provide the participant's with little benefit.

Changes occurred within some individual participants from training 1 to training 2 responses. Of the 11 participants given the additional training, 4 (15.4%) of the participants' MIT %correct scores improved. Children who never used a morpheme (4/11) (i.e., had omissions or phrasal modification errors) or used the morpheme in an alternating pattern with no regard for meaning (1/11) did not change their responses with additional practice. Of the 5 youngest participants in the sample, 4 of them did not use a morpheme. Improvements in morpheme accuracy were seen for participants who used prefixes and changed to suffixes (2/11), used the wrong suffix and changed to the correct suffix (3/11), or who marked or omitted inconsistently and began consistently marking (3/11). In some cases, these improvements were not reflected in their final score. Additional training items did not change overall scores; however, certain types of errors were reduced or eliminated with additional training items.

In summary, parent report of language learning ability was not significantly correlated to participants' performance on the MIT task but language exposure, age, and STM were significantly associated with MIT performance. When different error types were analyzed, these variables provided additional insights. Additional training items appeared to be helpful for some of the participants.

DISCUSSION

The purpose of this study was to evaluate if dynamic assessment, specifically a morpheme induction task (MIT), would be associated with other protocols previously shown to reliably index young bilingual children's language abilities. This would provide further support for the use of MIT when assessing bilingual children. It was expected that MIT performance would be significantly correlated to parent ratings of children's linguistic proficiencies and to children's performances on knowledge independent language processing tasks, such as nonword repetition (NWR) tasks. Surprisingly, parent ratings were not associated with children's MIT performances even when children were given partial credit for various off-target responses. In contrast, NWR was associated with MIT performance, as expected. Specifically, English NWR was associated with all three scoring schemes but Spanish NWR was only associated with %morphemed and %suffixed. One reason parent ratings and MIT may not have shown congruence in this study may have been because although the SLAS represents a valid measure for use with children with LI, it may not be sensitive in picking up differences among children with "low-average" to "high-average/above average" levels of language learning abilities. With a typically developing population, most children would be expected to receive "average" ratings across all domains of language. It is possible there was not enough variation in the group for the correlational design to work properly.

Additional variables were considered to examine their impact on MIT performance. It was predicted that age, language exposure, vocabulary, nonverbal IQ, family size, and birth order would not impact performance. As expected, vocabulary, nonverbal IQ, family size, and birth order were not associated with MIT performance. Unexpectedly, within the 5-6-year-old age range, age was significantly correlated with each of the MIT outcome measures and language exposure was also associated with MIT %correct performance. Boyer and Martin (2013) found that the youngest children in their study sample (3-4-year-olds) rarely used the target morpheme during testing. In the current study, the younger 5-year-olds also appeared to have the most difficulties providing the trained morpheme. Language exposure and MIT performance had a negative relationship, meaning that the less English the participants were exposed to in the home, the lower their performance on the MIT. However, when children's commission errors were credited, language exposure was no longer a significant predictor of performance. This indicates errors were not always an indication of poor learning abilities but in some cases might have indicated insufficient L2 exposure. This outcome was consistent with results from Kohnert and Danahy (2007), who found that language exposure impacted performance. They found that when the task was presented to children in their dominate language, performance increased, indicating that familiarity with the task language improved performance. In contrast, Boyer and Martin (2012) found that participant scores did not improve when L1 supports were incorporated into the task. Differences between the studies indicate that tasks in L1may free up metalinguistic skills, whereas L2 administration, even with a translation, stresses the language system enough to decrease performance.

In contrast to Roseberry and Connell who reported that almost all of their typically developing participants had successfully learned the task after 20 teaching items and that additional training items were not needed, the current study sample demonstrated high levels of variability in performance (*M*=79.12, *SD*=22.03, range: 42-100) with average scores falling well below the expected levels of proficiency. At the group level, additional training items did not consistently improve overall performance; however, some participants improved their accuracy with additional practice.

Because some of the variables were significantly correlated with MIT performance, a series of regressions were run to examine the relative predictive power of the variables. First, age was the only variable significantly predicting the variation of %morphemed scores. Younger participants made more errors with morpheme, often omitting or changing it. For %suffixed, age was still predicting this score but adding Spanish NWR and Best NWR explained additional variance. As participants developed stronger STM and phonologic skills in their L1, they began to hone their morphological skills. For %correct, English NWR was the only significant predictor for participants using the target morpheme. Different errors could be partially explained by differences among participants in their age, language exposure, and STM skills.

Errors produced by participants suggested a sequence of increasingly sophisticated off-target responses that can be captured by a 4-stage model of novel morpheme induction learning (depicted in Figure 2). Initially, young children are probably unaware of the novel morpheme's presence in the training sentences or its function. Thus, errors will consist of either unmodified responses ("*a ball"*) or phrasal modifications that mark the difference in children's responses (e.g., "*It's a part of a*

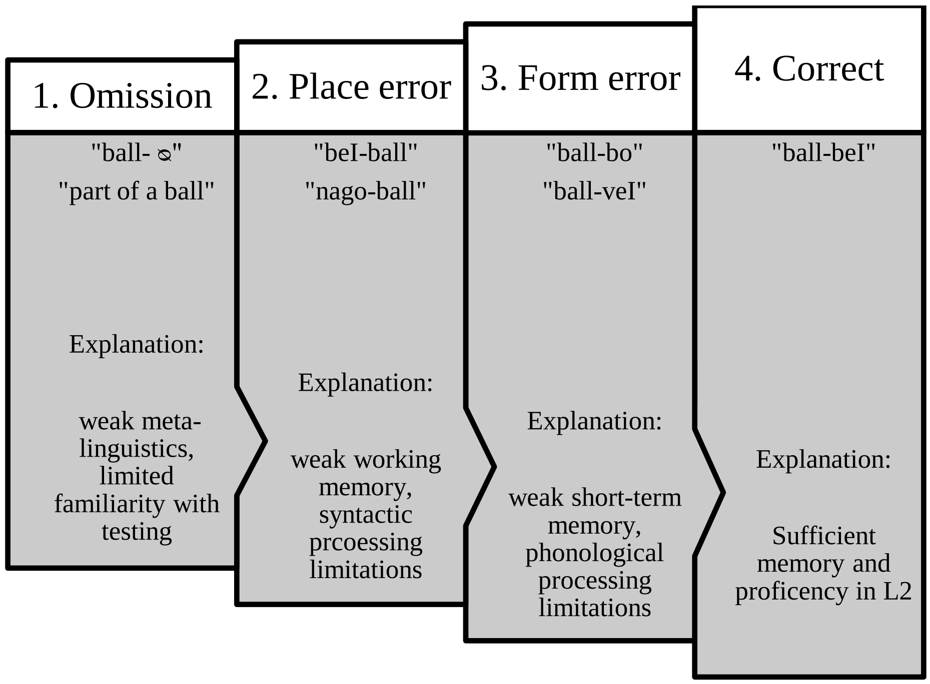

Figure 2. Suggested 4-stage model of morpheme induction learning.

ball"). These types of error were more common among the younger participants and suggest a limited understanding of the task, possibly due to underdeveloped metalinguistic skills or limited familiarity with general testing procedures. Next, children demonstrate detection of the novel morpheme's presence in the training item and its function by including a bound affix in their responses. However, responses included morphemes inaccurately placed in a prefix position (e.g., *"beI-ball*"). Weaknesses in processing nonnative English syntax or working memory could be responsible for the ability to mark the morpheme's existence and function but failure to register its position. In the next stage, children accurately place the morpheme in the suffix position of their responses but are still occasionally inaccurate with its phonological form, suggesting residual phonological weaknesses in completely accessing the morpheme's form. Stage 4 represents the correctly placed and fully fleshed-out version of the morpheme. A correct response indicates that not only has the new morpheme's position been properly registered in the context of the sentence but that its phonological form has also been accurately stored. Some support for the general sequence in the stage model was found in the changes that took place in those participants who improved with additional training. When they occurred, improved responses often reflected an adjustment to the next stage (e.g., moving from a place error response to a form error or from a form error to a correct response).

Limitations and Future Directions

The study has limitations that should be addressed in future investigations. First the study sample was relatively small, potentially making it underpowered. With only 26 participants, the study was only powered enough to detect moderate-to-large effect sizes. Small but potentially important effects could not be picked up in analyses. However, despite the small *N*, significant associations were still found. Future studies should replicate findings with a larger *N*.

Participant selection procedures also limit the interpretation of the results of the study. Children with low nonverbal abilities and hearing impairments were not included in the study. This limits the applicability of current findings to children with an IQ<80 and hearing impairments. Two potential participants were tested that had IQs below 80. Both participants performed poorly on the MIT and neither used morphemes in their responses. Before MIT is used with children with low IQs, additional studies are needed. Similarly, the age range was deliberately limited in the study so it would be a closer match to the age range in the Roseberry and Connell study. Results of the current study indicated that this particular MIT may work best for children who are older than 5;0 years.

Future studies should test directly the predictions of the proposed 4-stage model of morpheme induction learning by examining associations between variability in children's metalinguistic skills (e.g., grammaticality judgment tasks), working memory (e.g., backward digit span), and short-term memory skills (e.g., forward digit span) and their performance on the MIT. The model also makes the prediction that experimental manipulations that place stresses on different aspects of memory (e.g., divided attention task) during MIT learning should produce corresponding syntactic or phonological errors.

The present study aimed to include both typically developing language learners and children with low language abilities in proportions that reflect the population of English language learners from which they were drawn. The study sample consisted of primarily typically developing children but 3 of the 26 participants in the study were receiving speech and language services at the time of the study. Unfortunately, parental explanations of services were vague and exact diagnoses were not available for these participants. For these participants, performance varied across the MIT measure but all three performed poorly on English NWR. For details of their results, refer to Table 16. Additional study is needed to investigate further whether errors produced by children with language impairments during an MIT procedure are qualitatively different from those produced by children with typical development. If true, this would implicate differences between children with and without language impairments in the mechanisms they use to learn new morphemes.

Table 16. Descriptive statistics for group and children with SLP services.

	Group (*M* (*SD*) Range)	Child 1	Child 2	Child 3
Suspected Diagnosis	-	Speech sound disorder	Speech sound disorder and Language impairment	Late talker
Age (months)	70 (6.84) 60-83	81	66	62*
Years in school	1.9 (1.13) .5-5.5	4^	2	2
Spanish:English Exposure Ratio	.58 (.10) .31-.80	.31**	.80^^	.69^
Spanish:English Proficiency Ratio	.49 (.09) .31-67	.31**	.50	.67^
Mother's education	2.96 (1.216) 1-5 1Junior high-5 Master's/Doctorate	High school graduate	High school graduate	College graduate
Nonverbal abilities (NNAT)	107.58 (13.21) 86-133	128^	105	93*
NWR-English	75.40 (10.61) 59-94	61*	59*	63*
NWR-Spanish	80.27 (12.58) 35-94	58*	73	86
ROW-PVT:SBE standard score	103.36 (8.34) 90-123	109	96	98
Best SLAS composite	5.45 (1.14) 3.08-7.00	4.54	3.08	5.92
MIT %correct	79.12 (22.031) 42-100	100	50*	50*
MIT %marked	76.88 (37.327) 0-100	100	60	0**
MIT improvement	-	Not applicable	NO	NO
MIT errors	-	none	Omission, suffix error, prefix error (note many not included due to vocab pretest)	omission

*1 *SD* below mean **2 *SD* below mean
^1*SD* above mean ^^2*SD* above mean
NNAT= Naglieri Nonverbal Ability Test; NWR=Nonword repetition; ROWPVT= Receptive One-word Picture Vocabulary Test: Spanish-Bilingual Edition, 2012; SLAS= Speech and Language Assessment Scale; MIT= morpheme induction task

Concluding Remarks

The morpheme induction task developed more than 20 years ago by Roseberry and Connell represents a promising experimental procedure that could eventually be translated into routine clinical practice. It is one of the few procedures that can be validly administered by a monolingual clinician to a bilingual child suspected of having a language impairment. However, important gaps remain and additional investigations are needed before the full clinical potential of MIT can be realized.

REFERENCES

Acarlar, F., & Johnston, J. R. (2011). Acquisition of Turkish grammaticalmorphology by children with developemental disorders. *International Journal of Language & Communication Disorders*, *46*(6), 728-738.

Anderson, R. (2004). First language loss in Spanish-speaking children: Patterns of loss and implications for clinical practice. In B. A. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-speaking speakers* (pp. 187-212). Baltimore, MD: Brookes Publishing Company.

Bedore, L. M., & Leronard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminate function analysis. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1185-1192.

Bedore, L. M., & Leonard, L. B. (2000). The effects of inflectional variation on fast mapping of verbs in English and Spanish. *Journal of Speech, Language, and Hearing Research*, *43*, 21-33.

Bedore, L. M., & Leonard, L. B. (2005). Verb inflections and noun phrase morphology in the spontaneous speech of Spanish-speaking children with specific language impairment. *Applied Psycholinguistics*, *26*, 195-225.

Bedore, L. M., Pena, E. D., Joyner, D., & Macken, C. (2011). Parent and teacher rating of bilingual language proficiency and language development concerns. *International Journal of Bilingual Education and Bilingualism*, *14*(5), 489-511.

Bedore, L. M., Pena, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., . . . Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism: Language and Cognition*, *15*(3), 616-629.

Bishop, D. (2003). *The Children's Communication Checklist-2*. London: Psychological Publications.

Bishop, D. V., Laws, G., Adams, C., & Norbury, C. F. (2006). High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. *Behavior Genetics*, *36*(2), 173-184.

Bishop, D. V., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders*, *44*(5), 600-615.

Boudreau, D. (2005). Use of a parent questionnaire in emergent and early literacy assessment of preschool children. *Language, Speech and Hearing Services in Schools*, *36*, 33-47.

Boyer, V. E., & Martin, K. Y. (2012). Invented rule with English language learners. *Clinical Linguistics & Phonetics*, *26*(7), 613-627.

Brownell, R. (2012). *Receptive One-Word Picture Vocabulary Test: 2012 Edition-Spanish-Bilingual Edition*. Novato, CA: Academic Therapy Publications.

Caesar, L. G., & Kohler, P. D. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools*, *38,* 190-200.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language, and Hearing Research*, *40*, 519-525.

Connell, P. J. (1987). An effect of modeling and imitation teaching procedures on children with and without specific language impairment. *Journal of Speech and Hearing Research*, *30,* 105-113.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *Journal of Child Psychology and Psychiatry*, *42*, 741-748.

Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy.* Clevedon, Avon, England: Multilingual Matters.

Dale, P. S. (1991). The validity of a parent report measure on vocabulary and syntax at 24 months. *Journal of Speech and Hearing Research*, *34,* 565-571.

Diamond, K., & Squires, J. (1993). The role of parental report in the screening and assessment of young children. *Journal of Early Intervention*, *17*, 107-115.

Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1136-1146.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language and Hearing Research*, *54*, 1077-1088.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test- Revised*. Circle Pines, MN: American Guidance Service.

Ebert, K. D., Kalanek, J., Cordero, K. N., & Kohnert, K. (2008). Spanish nonword repetition: Stimuli development and preliminary results. *Communication Disorders Quarterly*, *29*(2), 67-74.

Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., . . . Reilly, J. (1993). *The MacArthur communicative development inventories*. San Diego, CA: Singular.

Gabor, B., & Lukacs, A. (2012). Early morphological productivity in Hungarian: Evidence from sentence repetition and elicited production. *Journal of Child Language*, *39*(2), 411-442.

Gathercole, C., & Baddeley, A. (1990). Phonological memory deficits in language disordered children: Is there a casual connection? *Journal of Memory and Language*, *29*, 336-360.

Gathecole, S. E., Willis, C., & Baddeley, A. D. (1991). Differentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. *British Journal of Psychology*, *82*, 387-406.

Gillam, R., Pena, E. D., & Miller, L. (1999). Dynamic assessment of narrative and expository discourse. *Topics in Language Disorders*, *20*(1), 33-47.

Goldman, R., & Fristoe, M. (1986). *The Goldman-Fristoe Test of Articulation*. Circle Pines, MN: American Guidance Service.

Gorman, B. K. (2012). Relationships between vocabulary size, working memory, and phonological awaremness in Spanish-speaking English language learners. *American Journal of Speech-Language Pathology*, *21*, 109-123.

Gray, S. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, *36*, 129-151.

Guiberson, M., Rodriguez, B. L., & Dale, P. S. (2011). Classification accuracy of brief parent report measures of language development in Spanish-speaking toddlers. *Language, Speech, and Hearing Services in the Schools*, *42*, 536-549.

Gutierrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher report. *Applied Psycholinguistics*, *24*, 267-288.

Gutierrez-Clellen, V. F., Kreiter, & Zagursky. (2007). *Ability and use of language in the home* [survey form].

Gutierrez-Clellen, V. F., & Pena, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech, and Hearing Services in Schools*, *32*, 212-224.

Hadley, P. A., & Rice, M. L. (1993). Parental judgments of preschoolers' speech and language development: A resource for assessment and IEP planning. *Seminars in Speech and Language*, *14*, 278-288.

Jackson-Maldonado, D., Bates, E., & Thal, D. (1992). *Fundacion MacArthur: Inventario del desarrollo de habilidades communicativas*. San Diego, CA: San Diego University.

Jackson-Maldonado, D., Thal, D., Marchman, V. A., Newton, T., Bates, E., & Guitierrez-Clellen, V. F. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, *20*(3), 523-549.

Jacobs, E. L., & Coufal, K. L. (2001). A computerized screening instrument of language learnability. *Communication Disorders Quarterly*, *22*, 67-75.

Kan, P. F., & Kohnert, K. (2008). Fast mapping by bilingual preschool children. *Journal of Child Language*, *35*, 495-514.

Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech and Hearing Services in Schools*, *43*, 81-96.

Klee, T., Carson, D. K., Gavin, W. J., Hall, L., Kent, A., & Reece, S. (1998). Concurrent and predictive validity of an early language screening program. *Journal of Speech, Language and Hearing Research*, *41*(3), 627-641.

Kohnert, K., & Danahy, K. (2007). Young L2 learners' performance on a novel morpheme task. *Clinical Linguistics and Phonetics*, *21*, 557-569.

Leonard, L. (1998). *Specific language impairment.* Cambridge, MA: The MIT Press.

Leonard, L. B., Sabbadini, L., Leonard, J., & Volterra, V. (1987). Specific language impairment in children: A cross-linguistic study. *Brain and Language*, *32*, 233-252.

Linder, T. W. (1993). *Transdisciplinary play-based assessment.* Baltimore, MD: Paul H. Brookes.

Lukacs, A., Laurence, L. B., & Kas, B. (2010). Use of noun morphology by children with language impairment: The case of Hungarian. *International Journal of Language & Communication Disorders*, *45*(2), 145-161.

Marchman, V. A., Martinez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, *7*(2) 212-224.

Massa, J., Gomes, H., Tartter, V., Wolfson, V., & Halperin, J. M. (2008). Concordance rates between parent and teacher clinical evaluation of language fundamentals observational rating scale. *International Journal of Language & Communication Disorders*, *43*(1), 99-110.

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Reexamining the age factor*. Amsterdam: John Benjamins.

Naglieri, J. A. (2004). *Naglieri Nonverbal Ability Test*. San Antonio, TX: Pearson Education.

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist-2. *International Journal of Language & Communication Disorders*, *39*(3), 345-364.

Paradis, J., Emmerzael, K., & Duncan, T. S. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders*, *43*, 474-497.

Paradis, J., Schneider, P., & Duncan, T. S. (2013). Discriminating children with language impairment amoung English-language learners from diverse first-language backgrounds. *Journal of Speech, Language, and Hearing Research*, *56*, 971-981.

Patterson, J. L., Rodriguez, B. L., & Dale, P. S. (2013). Response to dynamic language tasks among typically developing Latino preschool children with bilingual experience. *American Journal of Speech-Language Pathology*, *22*, 103-112.

Paul, R., & Norbury, C. F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing, and communicating* (4th ed.). St. Louis, MO: El Sevier Mosby.

Pena, E. D., Gillam, R. B., Malek, M., Ruiz-Felter, R., Resendiz, M., Fiestas, C., & Sabel, T. (2006). Dynamic assessment of school-age children's narrative ability: An experimental investigation of classification accuracy. *Journal of Speech, Language and Hearing Research*, *4*, 1037-1057.

Pena, E. D., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word-learning ability. *American Journal of Speech-Language Pathology, 10*, 138-154.

Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, *54*, 587-599.

Rescorla, L. (1993). Use of parental report in the identification of communicatively delayed toddlers. *Seminars in Speech and Language*, *14*, 264-277.

Rescorla, L., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language, and Hearing Research*, *44*, 434-445.

Restrepo, M. A. (1998). Identifiers of predominantly Spanish-speaking children with language impairment. *Journal of Speech, Language, and Hearing Research*, *41*, 1398-1411.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, *41*, 1412-1431.

Roseberry, C. A., & Connell, P. J. (1991). The use of an invented language rule in the differentiation of normal and language-impaired Spanish-speaking children. *Journal of Speech and Hearing Research*, *34*, 596-603.

Saenz, T. I., & Huer, M. B. (2003). Testing strategies involving least biased language assessment of bilingual children. *Communication Disorders Quarterly*, *24*(4), 184-193.

Schmid, M. S. (2011). *Language attrition.* New York: Cambridge University Press.

Sharp, K. M., & Gathercole, V. C. (2013). Can a novel work repetition task be a language-neutral assessment tool? Evidence from Welsh-English bilingual children. *Child Language Teaching and Therapy*, *29*(1), 77-89.

Skarakis-Doyle, E., Campbell, W., & Dempsey, L. (2009). Identification of children with language impairment: Investigating the classification accuracy of the MacArthur-Bates Communication Development Inventories, level III. *American Journal of Speech-Language Pathology*, *18*, 277-288.

Squires, J., Potter, L., & Bricker, D. (1999). *Ages and Stages Questionnaire user's guide*. Baltimore, MD: Brookes.

Summers, C., Bohman, T. M., Gillam, R. B., Pena, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language & Communication Disorders*, *45*(4), 480-493.

Thal, D., Jackson-Maldonado, D., & Acosta, D. (2000). Validity of parent-report measure of vocabulary and grammar for Spanish-speaking toddlers. *Journal of Speech, Language, and Hearing Research*, *43*, 1087-1110.

Thal, D. J., O'Hanlon, L., Clemmons, M., & Franlin, L. (1999). Validity of a parent report measure of vocabulary and syntax for preschool children with language impairment. *Journal of Speech, Language, and Hearing Research*, *42*, 482-496.

Thordardottir, E., & Brandker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*, 1-16.

Ukrainetz, T. A., Harpell, S., Walsh, C., & Coyle, C. (2000). A preliminary investigation of dynamic assessment with Native American kindergartners. *Language, Speech, and Hearing Services in Schools, 31*, 142-154.

United States Census Bureau. (2009). *Population: Ancestry, language spoken at home.* Retrieved from The National Data Book: http://www.census.gov/compendia/statab/cats/population/ancestry_language_spoken_at_home.html

Menlove, M., Gallo, G., & Arbogast, L. (2013). *Special education rules*. Salt Lake City, UT: Utah State Office of Education.

Vygotsky, L. (1980). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Weinberg, A. M. (1991). *Construct validity of the speech and language assessment scale: a tool for recording parent judgments.* Department of Speech-Language-Hearing, University of Kansas, Lawrence, KS.

Windsor, J., Kohnert, K., Lobitz, K., & Pham,G. (2010). Cross-Language nonword repetition by bilingual and monolingual children. *American Journal of Speech-Language Pathology*, *19*, 298-310.